

IBM_Data_Science_Professional_Certificate_Databases_and_SQL_for_Da

October 2, 2023

```
[1]: import csv,sqlite3
import pandas as pd
```

```
[2]: %load_ext sql
```

```
[5]: Connection=sqlite3.connect("Chicago_Improve_Educational.db")
Cursor=Connection.cursor()
```

```
[6]: %sql sqlite:///Chicago_Improve_Educational.db
```

```
[6]: 'Connected: @Chicago_Improve_Educational.db'
```

```
[7]: Socioeconomic_Indicators_in_Chicago=pd.read_csv("https://data.cityofchicago.org/
↪resource/kn9c-c2s2.csv")
Socioeconomic_Indicators_in_Chicago.head()
```

```
[7]:      ca community_area_name  percent_of_housing_crowded  \
0    1.0          Rogers Park                7.7
1    2.0          West Ridge                7.8
2    3.0           Uptown                3.8
3    4.0    Lincoln Square                3.4
4    5.0      North Center                0.3

      percent_households_below_poverty  percent_aged_16_unemployed  \
0                23.6                8.7
1                17.2                8.8
2                24.0                8.9
3                10.9                8.2
4                 7.5                5.2

      percent_aged_25_without_high_school_diploma  \
0                18.2
1                20.8
2                11.8
3                13.4
4                 4.5
```

	percent_aged_under_18_or_over_64	per_capita_income_	hardship_index
0	27.5	23939	39.0
1	38.5	23040	46.0
2	22.2	35787	20.0
3	25.5	37524	17.0
4	26.2	57123	6.0

```
[8]: Chicago_Public_Schools=pd.read_csv("https://data.cityofchicago.org/resource/↵9xs2-f89t.csv")
Chicago_Public_Schools.head()
```

```
[8]: school_id                                name_of_school \
0      610038                                Abraham Lincoln Elementary School
1      610281  Adam Clayton Powell Paideia Community Academy ...
2      610185                                Adlai E Stevenson Elementary School
3      609993                                Agustin Lara Elementary Academy
4      610513                                Air Force Academy High School

elementary_or_high_school    street_address    city state    zip_code \
0                          ES      615 W Kemper Pl    Chicago    IL      60614
1                          ES    7511 S South Shore Dr    Chicago    IL      60649
2                          ES      8010 S Kostner Ave    Chicago    IL      60652
3                          ES      4619 S Wolcott Ave    Chicago    IL      60609
4                          HS      3630 S Wells St    Chicago    IL      60609

phone_number                                link_ \
0 (773) 534-5720  http://schoolreports.cps.edu/SchoolProgressRep...
1 (773) 535-6650  http://schoolreports.cps.edu/SchoolProgressRep...
2 (773) 535-2280  http://schoolreports.cps.edu/SchoolProgressRep...
3 (773) 535-4389  http://schoolreports.cps.edu/SchoolProgressRep...
4 (773) 535-1590  http://schoolreports.cps.edu/SchoolProgressRep...

network_manager    ...    rcdts_code    x_coordinate \
0      Fullerton Elementary Network    ...    1500000000000000    1171699.458
1      Skyway Elementary Network    ...    1500000000000000    1196129.985
2      Midway Elementary Network    ...    1500000000000000    1148427.165
3      Pershing Elementary Network    ...    1500000000000000    1164504.290
4      Southwest Side High School Network    ...    1500000000000000    1175177.622

y_coordinate    latitude    longitude    community_area_number \
0    1915829.428    41.924497    -87.644522    7
1    1856209.466    41.760324    -87.556736    43
2    1851012.215    41.747111    -87.731702    70
3    1873959.199    41.809757    -87.672145    61
4    1880745.126    41.828146    -87.632794    34
```

	community_area_name	ward	police_district	location
0	LINCOLN PARK	43	18	(41.92449696, -87.64452163)
1	SOUTH SHORE	7	4	(41.76032435, -87.55673627)
2	ASHBURN	13	8	(41.74711093, -87.73170248)
3	NEW CITY	20	9	(41.8097569, -87.6721446)
4	ARMOUR SQUARE	11	9	(41.82814609, -87.63279369)

[5 rows x 79 columns]

```
[9]: Chicago_Crime_Data=pd.read_csv("https://data.cityofchicago.org/resource/
    ↪ijzp-q8t2.csv")
Chicago_Crime_Data.head()
```

```
[9]:      id case_number      date      block iucr \
0  13224000    JG440166  2023-09-23T00:00:00.000    067XX S PEORIA ST  0810
1  13227265    JG444151  2023-09-23T00:00:00.000    031XX N ORCHARD ST  0890
2  13220374    JG436141  2023-09-23T00:00:00.000    004XX N LAKE SHORE DR  1242
3  13220209    JG435819  2023-09-23T00:00:00.000    045XX W FULLERTON AVE  0910
4  13219931    JG435260  2023-09-23T00:00:00.000    061XX S WABASH AVE  0820
```

	primary_type	description	location_description	arrest	domestic	\
0	THEFT	OVER \$500	APARTMENT	False	False	
1	THEFT	FROM BUILDING	APARTMENT	False	False	
2	DECEPTIVE PRACTICE	COMPUTER FRAUD	APARTMENT	False	False	
3	MOTOR VEHICLE THEFT	AUTOMOBILE	STREET	False	False	
4	THEFT	\$500 AND UNDER	APARTMENT	False	True	

	ward	community_area	fbi_code	x_coordinate	y_coordinate	year	\
0	...	16	68	06	1171469.0	1860147.0	2023
1	...	44	6	06	NaN	NaN	2023
2	...	42	8	11	1180011.0	1903658.0	2023
3	...	31	20	07	1145827.0	1915574.0	2023
4	...	20	40	06	1177772.0	1864271.0	2023

	updated_on	latitude	longitude	\
0	2023-09-30T15:42:17.000	41.771705	-87.647004	
1	2023-09-30T15:42:17.000	NaN	NaN	
2	2023-09-30T15:42:17.000	41.890911	-87.614357	
3	2023-09-30T15:42:17.000	41.924327	-87.739595	
4	2023-09-30T15:42:17.000	41.782881	-87.623774	

	location
0	\n, \n(41.771704809, -87.647003596)
1	NaN
2	\n, \n(41.890910853, -87.614356731)
3	\n, \n(41.924326989, -87.739595106)
4	\n, \n(41.782881146, -87.623774252)

[5 rows x 22 columns]

```
[10]: Socioeconomic_Indicators_in_Chicago.  
      ↳to_sql("Socioeconomic_Chicago",Connection,if_exists="replace",method="multi",index=False)
```

[10]: 78

```
[11]: Chicago_Public_Schools.  
      ↳to_sql("Chicago_Schools",Connection,if_exists="replace",method="multi",index=False)
```

[11]: 566

```
[12]: Chicago_Crime_Data.  
      ↳to_sql("Chicago_Crime",Connection,if_exists="replace",method="multi",index=False)
```

[12]: 1000

```
[ ]: #Find the total number of crimes recorded in the CRIME table
```

```
[13]: %sql select count(*) from Chicago_Crime
```

```
sqlite:///Chicago_Improve  
* sqlite:///Chicago_Improve_Educational.db  
Done.
```

[13]: [(1000,)]

```
[ ]: #List community areas with per capita income less than 11000.
```

```
[15]: %sql select community_area_name, per_capita_income_ from Socioeconomic_Chicago  
      ↳where per_capita_income_<11000
```

```
sqlite:///Chicago_Improve  
* sqlite:///Chicago_Improve_Educational.db  
Done.
```

[15]: [('West Garfield Park', 10934),
 ('South Lawndale', 10402),
 ('Fuller Park', 10432),
 ('Riverdale', 8201)]

```
[ ]: #List all kidnapping crimes involving a child?
```

```
[30]: %sql select case_number,primary_type,description from Chicago_Crime where  
      ↳primary_type="KIDNAPPING"
```

```
sqlite:///Chicago_Improve  
* sqlite:///Chicago_Improve_Educational.db  
Done.
```

[30]: []

```
[ ]: #What kinds of crimes were recorded at schools?
```

```
[31]: %sql select primary_type,description,location_description from Chicago_Crime
      ↳where location_description like "%school%"
```

```
sqlite:///Chicago_Improve
* sqlite:///Chicago_Improve_Educational.db
Done.
```

```
[31]: [('DECEPTIVE PRACTICE', 'CREDIT CARD FRAUD', 'SCHOOL - PUBLIC BUILDING'),
      ('MOTOR VEHICLE THEFT', 'ATTEMPT - AUTOMOBILE', 'SCHOOL - PUBLIC GROUNDS'),
      ('ASSAULT', 'SIMPLE', 'SCHOOL - PRIVATE GROUNDS'),
      ('BATTERY', 'SIMPLE', 'SCHOOL - PUBLIC BUILDING'),
      ('BATTERY', 'SIMPLE', 'SCHOOL - PUBLIC BUILDING'),
      ('DECEPTIVE PRACTICE', 'FRAUD OR CONFIDENCE GAME', 'SCHOOL - PRIVATE GROUNDS'),
      ('CRIMINAL DAMAGE', 'TO PROPERTY', 'SCHOOL - PUBLIC GROUNDS'),
      ('DECEPTIVE PRACTICE', 'CREDIT CARD FRAUD', 'SCHOOL - PUBLIC BUILDING'),
      ('OFFENSE INVOLVING CHILDREN', 'CHILD ABUSE', 'SCHOOL - PRIVATE GROUNDS'),
      ('BATTERY', 'AGGRAVATED - OTHER DANGEROUS WEAPON', 'SCHOOL - PUBLIC GROUNDS'),
      ('BATTERY', 'SIMPLE', 'SCHOOL - PUBLIC GROUNDS')]
```

```
[ ]: #List the average safety score for each type of school.
```

```
[32]: %sql select elementary_or_high_school, avg(safety_score) from Chicago_Schools
      ↳group by elementary_or_high_school
```

```
sqlite:///Chicago_Improve
* sqlite:///Chicago_Improve_Educational.db
Done.
```

```
[32]: [('ES', 49.52038369304557), ('HS', 49.62352941176471), ('MS', 48.0)]
```

```
[ ]: #List 5 community areas with highest % of households below poverty line
```

```
[34]: %%sql
select community_area_name,percent_households_below_poverty from
      ↳Socioeconomic_Chicago
order by percent_households_below_poverty desc limit 5
```

```
sqlite:///Chicago_Improve
* sqlite:///Chicago_Improve_Educational.db
Done.
```

```
[34]: [('Riverdale', 56.5),
      ('Fuller Park', 51.2),
      ('Englewood', 46.6),
      ('North Lawndale', 43.1),
```

```
('East Garfield Park', 42.4)]
```

```
[ ]: #Which community area is most crime prone?
```

```
[36]: %%sql
select community_area, count(*) as crime_rate from Chicago_Crime group by
      ↪community_area order by crime_rate desc limit 1
```

```
sqlite:///Chicago_Improve
* sqlite:///Chicago_Improve_Educational.db
Done.
```

```
[36]: [(28, 44)]
```

```
[ ]: #Use a sub-query to find the name of the community area with highest hardship
      ↪index
```

```
[40]: %sql select community_area_name,(select max(hardship_index) from
      ↪Socioeconomic_Chicago) from Socioeconomic_Chicago
```

```
sqlite:///Chicago_Improve
* sqlite:///Chicago_Improve_Educational.db
Done.
```

```
[40]: [('Rogers Park', 98.0),
      ('West Ridge', 98.0),
      ('Uptown', 98.0),
      ('Lincoln Square', 98.0),
      ('North Center', 98.0),
      ('Lake View', 98.0),
      ('Lincoln Park', 98.0),
      ('Near North Side', 98.0),
      ('Edison Park', 98.0),
      ('Norwood Park', 98.0),
      ('Jefferson Park', 98.0),
      ('Forest Glen', 98.0),
      ('North Park', 98.0),
      ('Albany Park', 98.0),
      ('Portage Park', 98.0),
      ('Irving Park', 98.0),
      ('Dunning', 98.0),
      ('Montclair', 98.0),
      ('Belmont Cragin', 98.0),
      ('Hermosa', 98.0),
      ('Avondale', 98.0),
      ('Logan Square', 98.0),
      ('Humboldt park', 98.0),
      ('West Town', 98.0),
```

('Austin', 98.0),
('West Garfield Park', 98.0),
('East Garfield Park', 98.0),
('Near West Side', 98.0),
('North Lawndale', 98.0),
('South Lawndale', 98.0),
('Lower West Side', 98.0),
('Loop', 98.0),
('Near South Side', 98.0),
('Armour Square', 98.0),
('Douglas', 98.0),
('Oakland', 98.0),
('Fuller Park', 98.0),
('Grand Boulevard', 98.0),
('Kenwood', 98.0),
('Washington Park', 98.0),
('Hyde Park', 98.0),
('Woodlawn', 98.0),
('South Shore', 98.0),
('Chatham', 98.0),
('Avalon Park', 98.0),
('South Chicago', 98.0),
('Burnside', 98.0),
('Calumet Heights', 98.0),
('Roseland', 98.0),
('Pullman', 98.0),
('South Deering', 98.0),
('East Side', 98.0),
('West Pullman', 98.0),
('Riverdale', 98.0),
('Hegewisch', 98.0),
('Garfield Ridge', 98.0),
('Archer Heights', 98.0),
('Brighton Park', 98.0),
('McKinley Park', 98.0),
('Bridgeport', 98.0),
('New City', 98.0),
('West Elsdon', 98.0),
('Gage Park', 98.0),
('Clearing', 98.0),
('West Lawn', 98.0),
('Chicago Lawn', 98.0),
('West Englewood', 98.0),
('Englewood', 98.0),
('Greater Grand Crossing', 98.0),
('Ashburn', 98.0),
('Auburn Gresham', 98.0),

```
('Beverly', 98.0),  
( 'Washington Height', 98.0),  
( 'Mount Greenwood', 98.0),  
( 'Morgan Park', 98.0),  
( "O'Hare", 98.0),  
( 'Edgewater', 98.0),  
( 'CHICAGO', 98.0)]
```

```
[ ]: #Use a sub-query to determine the Community Area Name with most number of  
      → crimes?
```

```
[41]: %sql select community_area, count(*) as crime_rate from Chicago_Crime group by  
      → community_area order by crime_rate desc limit 1
```

```
sqlite:///Chicago_Improve  
* sqlite:///Chicago_Improve_Educational.db  
Done.
```

```
[41]: [(28, 44)]
```